Advances in Comparative Assessment

Nijkamp, P.

1999

ADVANCES IN COMPARATIVE ASSESSMENT RESEARCH IN THE SPACE-ECONOMY

Peter Nijkamp

Research Memorandum 1999 - 5

vrije Universiteit amsterdam

ADVANCES IN COMPARATIVE ASSESSMENT RESEARCH IN THE SPACE-ECONOMY

Peter Nijkamp

Research Memorandum 1999 - 5

ADVANCES IN COMPARATIVE ASSESSMENT RESEARCH IN THE SPACE-ECONOMY

Peter Nij kamp

Abstract

This paper aims to give a survey of recent assessment methods in spatial planning. The attention is focused on three classes of evaluation methods, viz. regime analysis, the flag model and rough set analysis. The various methods are discussed and their strengths and weaknesses are mutually compared. The paper is concluded with some retrospective and prospective remarks.

1. The Changing Scene of Assessment

Decision-making in a complex space-economy is fraught with many difficulties, as is witnessed by an abundance of cases in regional policy, transport planning and environmental management. These difficulties can partly be ascribed to lack of information (imprecise and incomplete data, insufficient knowledge base on long-range issues etc.), but also partly to the multidimensional conflict nature of modern regional, urban, transportation and environmental issues. Such conflicts may concern the trade-offs between different policy objectives, different interests among stakeholders in the space-economy, different competencies of decision levels, etc. (see also Nijkamp et al. 1992).

Conflict resolution is of course a political action, but presupposes proper knowledge on the pros and cons of alternative choice possibilities. From an economic perspective, this would imply that all foreseeable costs and benefits of a planned initiative would have to be assessed. But in a broader context, also the social, cultural, environmental and safety aspects would have to be considered, as is clearly witnessed by e.g. plans to build a Trans European Network (TEN) or to develop sustainable tourism areas.

Policy analysis offers an assessment and evaluation framework in the public sector with the goal to increase the efficiency and effectiveness of government decisions. In this sector in particular, a wide range of decisions is to be made without a clear reliance on the market system. This is partly caused by the nature of choices in the public sector (with emphasis on multi-actor democratic modes of decision-making) and partly by the complexity of government projects (with long-lasting and often uncertain implications). And it is indeed increasingly recognised that decisions based on market forces alone do not necessarily lead to optimal results. Structural market failures as well as unexpected external factors may require an efficient policy mechanism that is able to lay the foundation for an improvement of the actual socio-economic developments within a community or society (Rietveld and Bruinsma 1998). Clearly, the initiation of structural policies or the implementation of corrective measures is often not the responsibility of a single government agency, but rather may take place on several organisational levels ranging from local to supranational.

In the past decades several methods have been developed and applied in policy analysis, in which a market evaluation played a prominent role. The most well-known example of such a market evaluation method is based on **cost-benefit analysis** (as an operational application of welfare theory). This method forms the foundation for many policy assessment methods and has been successfully applied in many case studies. Despite its great many merits, it is increasingly recognised in modem policy analysis that it also has some limitations, because not all relevant welfare implications of transport initiatives can be expressed in the 'measurement rod of money'. Therefore, a concise review of the potential of cost-benefit analysis may be helpful.

The foundation of cost-benefit analysis rests on solid economic grounds when it addresses the question: does the government (or a decision-making body) receive value for money? In other words, cost-benefit analysis assesses the benefits of a project in the light of the underlying costs. This is undoubtedly a healthy computational principle, and belongs to the standard economic toolbox. Many studies in the regional, transport and environmental field have followed these principles and have led to relevant results. Clearly, in practice a great variety of problems have emerged such as: the definition of benefits (e.g., based on willingness-to-pay principles), the distribution of costs and benefits (who pays, who benefits), the social rate of discount, the demarcation of the relevant area of the project, the existence of externalities, etc. A vast amount of literature has been published to address these questions. Theoretically, much progress has been made to cope with the limitations of conventional cost-benefit analysis, but in practice the results of such studies were not easily accepted by policy-makers, as is witnessed by discussions on cost-benefit studies for labour market policies or military defence projects. As a consequence, the use of cost-benefit studies in various project evaluation problems has sometimes led to debatable outcomes which did not receive sufficient scientific and political support. For example, in the transportation field we have observed the application of many cost-benefit studies, in particular for infrastructure projects. But also here recent studies (e.g., on the Channel Tunnel, the Dutch Betuwelijn, the Nordic Scanlink connections) have met with criticism, mainly because the underlying economic assumptions were feeble or because negative spill-over effects and social costs were not adequately quantified and included in the analysis.

We may conclude that cost-benefit studies seem to be most applicable and appropriate if the decision concerns a well demarcated and a priori precisely defined project which does not generate many unpriced externalities. If however, the decision concerns a more general policy programme (of which the details and even sometimes the major features are unknown), then the translation of its impacts into precisely measurable and quantitative consequences and subsequently into monetary figures is often rather problematic. Similarly, if a public investment is likely to generate a wide diversity of social costs (e.g., landscape destruction, loss of safety, health effects, loss of biodiversity or rare species, destruction of archaeological sites), it is often a heroic research task to come up with reliable figures which are broadly accepted in the policy arena. This does not mean that cost-benefit analysis would have to be discredited; but it would have to be complemented with more appropriate evaluation tools (Sikow-Magny and Niskanen 1998).

Thus, the (positive and negative) features of cost-benefit analysis are well-known and generally accepted, not only in neo-classical welfare economics but also in decision-making procedures which incorporate socio-economic aspects (see Janssen 1991). As a further enrichment of the cost-benefit analysis, modem multi-dimensional assessment approaches aim to merge and feature the different aspects which intervene as pros and cons during a decision-making process in the public sector.

A great diversity of modem assessment methods has been developed over the last ten years to extend the range of and to provide a complement to conventional cost-benefit analysis and to offer a perspective for procedural types of decision-making in which various qualitative aspects are also incorporated. Many of these methods simultaneously investigate the impacts of policy strategies on a multitude of relevant criteria, partly monetary, partly non-monetary (including qualitative facets). They are often coined **multicriteria methods** and are also known as multi-assessment methods.

It is noteworthy that in past years an avalanche of assessment studies has been undertaken in the regional, transportation and environmental field, but an integral study and a systematic comparison of findings of previously undertaken assessment studies is often difficult due to different analytical approaches and differences in presentation. The gradual shift from conventional assessment techniques (such as cost-benefit analysis) towards multi-dimensional assessment approaches (such as multicriteria analysis) has prompted the need for a systematic comparison of these studies, but this requires an enormous study effort and induces, as a consequence, a significant research cost.

Fortunately, over the past two decades a new set of research techniques has been developed which makes a rigorous analysis of study findings possible, **viz. meta-analysis.** Meta-analysis aims to summarise results from previous studies in a (preferably) quantitative way so as to also allow for transferability of findings (see, for details Van den Bergh et al. 1997). As a result of a more rigorous statistical - or at least analytical - underpinning, a synthesising process becomes more manageable and less vulnerable to subjective elements due to a more systematic investigation of early research findings. For example, voluminous study results can be analysed and the impact of ad hoc approaches on study findings can be reduced via the use of comparative, often quantitative methods which allow for a rigorous synthesis. These recent scientific developments make it possible to establish a new type of assessment methodology in order to address multi-dimensional decision problems in a rigorous way which could lead to a significant cost reduction due to the use of previously obtained knowledge.

Against the above sketched background, the aim of the present paper to offer a meta-analytical contribution to assessment methodology by comparing three rather recently developed assessment methods which have a great potential for evaluating and comparing different courses of action in the space-economy. These methods are: the **regime method** (a multicriteria method suitable for both quantitative and qualitative evaluation problems), the **flag model** (a method that is particularly appropriate for selecting the most suitable course of action in case of critical threshold values) and **rough set analysis** (a recently developed classification method which is particularly suitable for comparing in a qualitative sense various choice options). After a further exposition of assessment principles (section 2), these methods will successively be described in Sections 3 to 5. After some illustrations, the paper will offer also some concluding remarks.

2. Assessment in Policy Analysis

Policy analysis is the scientific preparation for decision-making. This is normally a procedural activity with many steps before an ultimate result is achieved. In general, we define the decision process as a set of actions and dynamic factors (behavioural, contextual) which, after a description of the problem and its alternative solutions for action, leads into a specific commitment to action (Janssen 1991). A first important element in this process is the identification of the problem which does not automatically exist, unless someone perceives it as such. For example, Ackoff (198 1) observes that an individual or a group can perceive a problem if, in a given choice situation, there is a difference between the present state and the desired one. This can happen when *"(I) the individual or group has alternative courses of action available; (2) the choice of action can have a significant effect on this perceived difference; and (3) the individual or group is uncertain a priori as to which alternative should be selected"*.

Next, the problem will pass through the decision-making process where, directly or indirectly, decision-makers will provide judgments for reaching the best alternative solution to the problem. Following Simon's definition (1960) of the different phases which characterise a decision-making process, Mintzberg (1979) describes such a process as a trichotomic structure based upon the three prominent steps of **identification, development** and **selection.**

Identification consists of two parts: **recognition,** where the problem is identified and **diagnosis,** in which the cause-effect of the relationships among the attributes of the problem is determined. In the development phase, the routine called **search** tries to find ready-made solutions, whereas the routine **design** develops custom-made solutions. Finally, in the selection phase there are some specific routines. First, the routine **screen** that checks which of the ready-made solutions are suitable for the problem. Second, the **evaluation-choice** routine, which is divided into three steps: **judgment,** where each decision-maker decides according to his own procedure; **bargaining,** in which there is the selection among judgments which may be conflicting due to the conditions of various objectives; and **analysis,** which is the evaluation of the solution carried out by analysts.

Often there is confusion between assessment methods that are directed toward policy assessment and those directed toward project appraisal. A policy is a qualitative course of action, while a project is a demarcated action at a concrete level of implementation. Between the policy decision and project implementation we can define a sequential process in which the assessment methodology is designed to correspond step-wise to the needs of decision-makers. For example, if we have to assess a transport project, the data the assessment method will use or examine will mainly be **quantitative** data that necessarily or ideally express (from a micro point-of-view) the complexity of the project by considering such variables as the capacity of the infrastructure, transport cost, travel time, level of pollution, congestion, and the like.

On the other hand, when we encounter the problem of assessing a policy program (such as the feasibility of a TEN), we first have to consider that the decision-makers formulate the policy in **qualitative** terms, for example, a proposal that promotes transport safety, or a policy that strengthens economic and social cohesion in the EU. Therefore, the necessity emerges then to step back somewhat and examine precisely how the decision-makers define a given or proposed policy. In order to assess policy, decision-makers require proper analytical and informational tools, a fundamental feature of which is the act of communicating to the broader public. Creation of support through scientific assessment and information provision is then critical. Clearly, the difference between a policy and a project may be a gradual one. For instance, when a policy has been formulated on CO_2 reduction, at the end it ought to be specified how much CO_2 would have to be reduced and what such a reduction would mean in implementing transport projects. Thus, we are essentially talking here about two ends of a broad spectrum, from global to specific, and from qualitative to quantitative.

In order to reach a satisfactory policy in a complex environment, a careful process of decision-making is required which takes time and can be costly. The problems underlying a decision-making process in a spatial context may be subdivided into the following components:

- the information or data available always contain a component of uncertainty;
- the data or information may be stored in different data bases that may be difficult to access, manipulate, compare and study;
- a large set of - often conflicting - objectives or targets has to be taken into account;
- the decision-making process itself might be influenced by power relations or selfish motivations;
- a decision-making process has to take place within the shortest time possible to avoid countervailing effects.

This means that in any societal setting the best alternative or policy has to be determined which may boost public acceptability or at least social feasibility; in other words, the basic question is: what is the optimal policy? Theoretically, a decision-maker has to deal with an optimisation procedure, where from a set of alternatives the possible optimal choice is to be found, given the objectives and underlying conditions and constraints in real life, The previous steps and considerations can also be incorporated in a systematic way in the so-called **policy analysis temple** (see Figure 1), which summarizes in a succinct ways most of the previous observations.

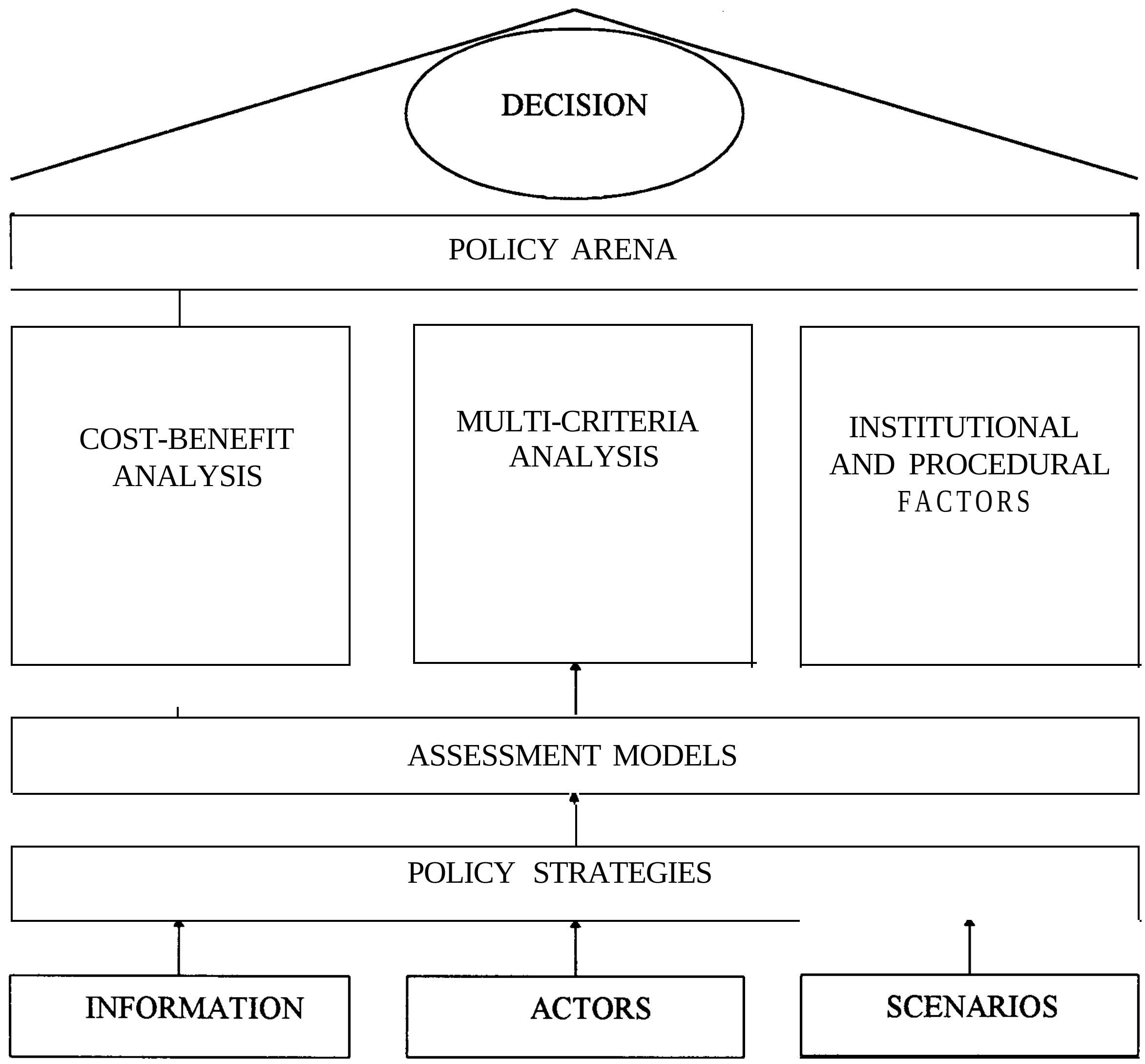


Figure 1. The Policy Analysis Temple

It goes without saying that uncertainty is a basic feature of policy analysis. Making decisions based on uncertain or imprecise information is a problem which has attracted the attention of many scientists; see, for example, Leung (1997) or Nijkamp and Scholten (1993). A wide range of support systems - which handle incomplete knowledge concerning real world phenomena - is now available, e.g. Decision Support Systems, Computer Information Systems and Expert Systems (see, for example, Jackson 1990). According to Kacprzyk and Yager (1990), these systems are built upon mathematical research techniques and aim to yield new knowledge via a proper treatment of data and/or information. However, in many situations uncertainty is not the only complicating factor in the decision-making process.

Most decisions can be typified as being of a multiple objective or multicriteria type (Janssen 1991, Nijkamp and Pepping 1998, Nijkamp and Blaas 1995). This means that an optimal alternative from a set of alternatives is to be determined which best satisfies a number of - often conflicting - objectives. Another complicating factor is that on the policy level - besides a set of quantitative criteria - qualitative criteria also must be taken into account in a decision-making process. Examples are the interest of the biotic and a-biotic environment, the protection of school children, accessibility conditions of the elderly generation, or the risk of criminality in public transport. In the past, the research has often resorted to cost-benefit analysis as on the appraisal method, and this has often been done in a successful way. However, as mentioned above, this method has severe shortcomings when it comes to an operationalisation of intangible facets, so that there are many justified reasons for the sometimes limited applicability of this method. In public policy evaluation, especially the study of environmental impacts turned out to be troublesome, since all advantages and disadvantages of policy options have to be translated into a common monetary unit. Hence, qualitative criteria of an unpriced and intangible nature cannot be included in the decision-making procedure based on a standard cost-benefit analysis. Within this approach, the market priorities are reflected in the (corrected) market prices or through the willingness-to-pay of the individuals (see Janssen 1991). In the practice of cost-benefit analysis, it was difficult to include incommensurable aspects of a project. Similarly, in the current practice in many countries there was hardly any applicable and meaningful way of including distributional impacts on welfare (e.g., through a weighting system for different groups) into policy evaluation, even though there is in the history of cost-benefit analysis theory in economic research a vast amount of literature of distributional issues (e.g. through weighting systems, social rates of discount, etc). Clearly, a complementary decision-making process better able to handle qualitative information in a more sophisticated way seems to be very useful with respect to decision-making.

The various interlinked steps can thus be represented systematically in the so-called policy analysis temple (see Figure 1). The common feature of modem multicriteria methods is that they are all able to include and to digest multiple conflicting criteria or objectives at both the individual level and the group (or community) level. For example, these methods are able to treat incommensurable decision variables, such as economic growth, social justice or environmental quality. In the literature we find a rich variety of multi-assessment methods (usually under the name of multicriteria methods or multiple objective evaluation methods). By setting aside the different labels, we can observe that these methods have one common element: the existence of multiple judgement (or evaluation) criteria. By considering a classification into discrete and continuous decision models, we can make a general distinction among multi-assessment methods. On the one hand, the discrete multiple assessment methods are structured to examine a finite number of feasible choices (e.g., in infrastructure evaluation). The continuous multiple assessment methods consider, on the other hand, an infinite number of choices in the decision-making process (e.g., in transport network models). The difference between these two classes is that multicriteria methods deal with a finite set of alternative choice possibilities (e.g., a distinct set of transport infrastructure investments), whereas multiple objective methods focus on infinite (essentially continuous) choice possibilities (e.g., the volumes of flows in a transport system or the level of CO_2 - emissions). Clearly, the mathematical treatment of these classes is different, but the underlying principle of policy analysis is the same.

The great variety of evaluation methods has confronted regional, transport and environmental researchers with intriguing choice problems. There is clearly not an unambiguous choice for a single, comprehensive evaluation method. The following stepwise approach seems to be feasible and justifiable:

- if all effects are quantitative and financial-monetary in nature, apply cost-benefit analysis.
- if all effects are quantitative, but not financial-monetary in nature, apply either a cost-effectiveness analysis ('value for money') or a quantitative assessment method (such as a weighted summation method or a goals achievement method).
- if (some or all) effects are qualitative in nature (and hence not transferable into the common measuring rod of money), apply a multicriteria method (see also Beinat and Nijkamp 1998).

Clearly, in the latter case we envisage a selection problem, viz. which method to employ? As argued extensively in Janssen (1991) and Nijkamp and Blaas (1995), in that case a practical approach to be followed is:

- identify the nature and the features of the specific evaluation problem under consideration.
- identify from the set of available multicriteria evaluation methods one method (or a class of methods) that best complies with the features of the choice problem (in terms of level of information, availability of weights, interactive decision-making procedure, etc).
- apply the latter method (or class of methods) to the evaluation problem at hand and find the final solution.

It goes without saying that in reality many real-world evaluation problems are characterised by imprecise, uncertain, fuzzy or sometimes only qualitative information. In that case one has to resort to multicriteria analysis.

In the framework of a strategic assessment of public policies the class of multicriteria methods seems to be very relevant as a decision support tool. In the past years, a broad range of multicriteria methods has been developed, applied and presented in the policy-analytical literature. It would be going too far to offer a complete overview, since this could easily amount to some hundred different methods (see also Jassen 1991). But there are certain main classes which deserve attention and will be presented in more detail in this study.

A first major distinction of multicriteria analysis concerns the **level of measurement,** viz. qualitative (i.e., categorical, nominal, binary, ordinal) versus quantitative (interval or ratio) information. The lower level of information in qualitative data imposes the use of specific analytical tools, e.g. based on pair-wise comparison, frequency analysis, or multidimensional scaling. Quantitative data can more easily be treated by means of the conventional rules of numerical analysis.

A second distinction concerns the difference between the identification of an optimal alternative versus the assignment of choice alternatives into distinct achievement classes. Thus, there is essentially a question of optimal **choice vis-à-vis a classification.**

A third distinction emerges from the question whether an absolute ranking of choice options is strived for, or whether a relative evaluation is given in relation to a reference situation (or ideal option). This is essentially a matter of **relative versus absolute assessment.**

And finally, some multicriteria methods serve to visualise only the characteristic differences between the attributes of choice alternatives, whereas others try to identify optimal solutions. Thus, here the main difference lies in **optimisation versus visualisation.**

Against these background remarks, in the sequel of this study we will present a limited, but interesting set of evaluation tools which comply with the above-mentioned four classification principles, viz:

- measurement level, i.e. qualitative vs. quantitative
- choice vs. classification
- relative vs. absolute assessment
- optimisation vs. visualisation

The methods which we will discuss here successively are:

- regime analysis
- flag model
- rough set analysis

Clearly, this class of multicriteria methods only becomes a relevant approach if traditional evaluation methods such as cost-benefit or cost-effectiveness analysis cannot be applied due to information shortage or specific requirements in a decision-support environment. Each of these methods has its strong and weak points and will concisely be described in subsequent sections.

3. The Multicriteria Regime Method

There is a vast array of discrete multicriteria evaluation methods ranging from simple weighted summation methods to complicated data correspondence techniques (see, for an overview Nijkamp et al. 1992). Some of these discrete assessment methods are mainly suitable for quantitative data, others are appropriate for qualitative data. Since many policy proposals have often (partly) qualitative data, we focus here on regime analysis. This multi-assessment method is a discrete evaluation method which has shown its suitability for a wide range of policy applications; its strengths are in particular: flexibility in assessing both projects and policies and its capability to analyse quantitative as well as ordinal and qualitative data.

The basic framework of standard multi-criteria methods is based upon two kinds of input data: an **evaluation** (on information, effect or impact) **matrix** and a set of **political weights.** The evaluation matrix is composed of elements which measure the effect of each relevant alternative in relation to each relevant criterion (attribute, feature or characteristic). The set of weights gives us information concerning the relative importance of the criteria under consideration. Regime analysis is a discrete multiple criteria method, and it may be conceived of as a generalised case of the widely used **concordance analysis.** Regime analysis is thus based on pair-wise comparison methods which are able to examine quantitative as well as qualitative data. In order to gain a better understanding of regime analysis, we will briefly reiterate the basic components of concordance analysis.

The concordance analysis is an evaluation method in which the basic idea is to rank a set of alternatives (choice possibilities, options, plans or proposals) on the basis of their pair-wise comparison in relation to relevant decision criteria. For instance, if we consider a choice problem where we have a set of alternatives and a set of criteria, then we begin our analysis by comparing alternative i with alternative j in relation to all criteria. After having done this, we select all criteria for which alternative i performs better than - or equal to - alternative j. This class of criteria is called a ‘concordance set’. Similarly, we define the class of criteria for which alternative i performs worse than alternative j. This set of criteria is called the ‘discordance set’.

We now need to rank the alternatives. In order to do so, we introduce the concordance index. The concordance index indicates the relative dominance of the alternatives in the concordance set. It is defined as the sum of the weights of the criteria according to which alternative i is more attractive than alternative j. Clearly, the higher the value of the concordance index an alternative, the more attractive this alternative is compared to others. Next, we may also define a discordance index which indicates the maximum difference of scores for the alternatives under consideration. When we seek for the best alternatives as a solution for our choice problem, we must select those alternatives that have the highest values for the concordance indices and the lowest values for the discordance indices.

The strength of regime analysis - based on the principles of the concordance analysis - is that it is able to deal with binary, ordinal, categorical and cardinal (ratio and interval) data, while it is also possible to use mixed data (i.e., partly qualitative, partly quantitative). This applies to both the effects and the weights in the policy analysis concerned. In a regime analysis - like in the concordance analysis - we compare the alternatives in relation to all relevant criteria in order to define the concordance index. Let us consider, for example, the comparison between alternative i and j. The concordance index is the sum of the weights which are related to the criteria for which i is better than j. Let us call this sum cij. Then we also calculate the concordance index for the same pair of alternatives, but now inversely, viz. by considering the criteria for which j is better than i, i.e., cji. After having calculated these two sums, we subtract these two values in order to obtain the net concordance index: $\mu_{ij}=c_{ij}-c_{ji}$.

If we have only ordinal information about the weights, our interest is focused on the sign of the index μ_{ij}. If the sign is positive, this will indicate that alternative i is more attractive than alternative j; if negative, it will imply the opposite. We will then in any case be able to rank our alternatives. We note that due to the relative nature of the information in the indicator μ, no value can be attached to the size of the difference between the alternatives; it is only the sign of the difference that is important.

Regime analysis is now very powerful in that it may also solve the complication that we may not be able to derive an unambiguous result, i.e. to rank all alternatives. This may happen in case of ambiguity in the sign of the index μ. In order to solve this problem, regime analysis introduces a (cardinal) probability pij for the dominance of criteria i with respect to criteria j as follows:

$$p_{ij} = prob \quad (\mu_{ij} > 0)$$

and we define an aggregate quasi-probability measure which indicates the success score as follows:

$$p_i = \frac{1}{I-1}\sum_{j\neq i} p_{ij}$$

where I is the number of alternatives.

The analytical problem now is to assess the value of pij (and of pi). The regime method then assumes a specific probability distribution of the set of a priori given, feasible weights. This assumption is based upon the criterion of Laplace in the case of decision-making under uncertainty. In the case of a rectangular probability distribution of qualitative information, it is sufficient to use a set of random stochastic drawings based on stochastic analysis, which is consistent with an originally ordinal data set. This procedure helps to overcome the methodological problem we can encounter by trying a numerical operation on qualitative data. Further technical details can be found in Nijkamp et al. (1992). From the viewpoint of numerical analysis, the regime method then identifies the feasible area in which values of the feasible weights wi must fall in order to be compatible with the condition imposed by their probability value. By means of a random generator, numerous values of weights can be calculated. This allows us at the end to calculate the performance score (or success score) pi for each alternative i. We can then determine an unambiguous - and even cardinally expressed - solution and rank order for the alternatives under consideration.

We can find many applications of the regime analysis in the literature on regional, environmental, land use and transportation planning (see Nijkamp et al. 1992, Nijkamp and Blaas 1995). These applications concern both project appraisal and policy assessment procedures. In particular, in the various case studies on policy assessment issues and processes, we find that the regime method is able to merge all different aspects of public decision-making problems into a systematic framework, which normally leads to unambiguous results with a cardinal meaning. Empirical examples include inter alia the evaluation of various trajectories of road projects, the choice between investments in private and public transport infrastructure, the location of new airports, etc. These examples include quantitative, qualitative and mixed information.

4. The Multicriteria Flag Model

Multicriteria analysis comprises a set of various multidimensional assessment and evaluation models. The flag model is a methodology that has recently been developed to offer a broad framework for decision support for sustainable development policy in the case of public decision-making, i.e. on land use or environmental investments (see Hermanides and Nijkamp 1998). A major issue in sustainability policy is how to determine a normative definition of sustainability. The objective of the flag model is to operationalise the concept of sustainability by defining a multicriteria approach in which the indicators are represented through ranges of values by using the normative concept of critical threshold values (see for a detailed application Nijkamp and Ouwersloot 1998).

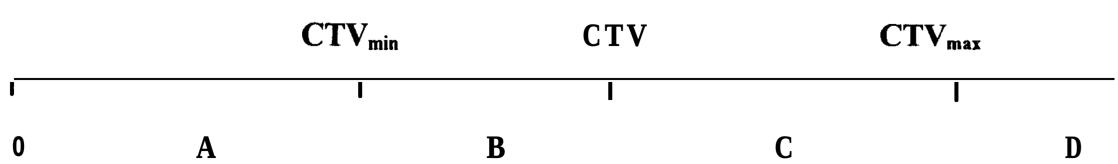
The flag model is a normative evaluation approach, which requires a multidimensional framework of analysis and of expert judgement which should be able to test actual and future states of the economy and the ecology against a set of normative reference values on sustainability. The flag model has been developed to assess the degree of sustainability of compound values of policy alternatives. The model develops an operational description and definition of the concept of sustainable development. There are three important components of the model:

1. identification of a set of measurable sustainability indicators;
2. establishment of a set of normative reference values;
3. development of a practical methodology for assessing future developments.

As mentioned above, the input of the programme is an impact matrix with a number of relevant policy variables or criteria; the matrix is formed by the values that the variables assume for each considered alternative (plan, scenario, etc.). Such values are defined by independent and qualified experts. The main purpose of the model is to analyse whether one or more alternatives can be classified as sustainable or not; such an evaluation is based upon the critical assessment indicators. The methodology therefore requires the identification and definition of policy relevant indicators (OECD 1993), which are suitable for further empirical treatment in the assessment procedure.

In the first place, the choice of indicators corresponds to the problem addressed; in general, the indicators must expose the problem under scrutiny as well as consider the objectives that such a problem must tackle. One significant dilemma we could encounter when defining the indicators is the likelihood that the number of indicators always tends to grow; and, to complicate matters, some indicators are encompassed within other indicators. In order to avoid the complication of a large number of indicators which would thus be difficult to examine and which are often minor and unnecessary, a helpful methodology is to use a **hierarchical** approach based on a tree-like structure. Such an approach corresponds to the idea of aggregation and disaggregation of the indicators that we deem fundamental to our examination. For instance, we can make distinctions among macro, meso and micro indicators, or distinguish by means of relevant time or geographical scales. Such indicators in the programme have two formal attributes: class and type.

Secondly, for each sustainable indicator we have to define the corresponding critical threshold values (CTVs) (see Figure 2). These values represent the reference system for judging actual states or future outcomes of scenario experiments. Since in certain areas and under certain circumstances experts and decision-makers may have conflicting views on the precise level of the acceptable threshold values, we may have to estimate a band width of values of the thresholds ranging from a maximum value (CTV_{max}) to a minimum value (CTV_{min}). This can be represented as follows:



Section **A**	**Green Flag**	no reason for specific concern
Section **B**	**Orange Flag**	be very alert
Section C	**Red Flag**	reverse trends
Section **D**	**Black Flag**	stop further growth

Figure 5. Critical Threshold Values (CTV) for policy variables

Finally, the third component of the model, the impact assessment, provides a number of instruments for the analysis of the sustainability issue. Each policy option has - as in any multicriteria analysis - a range of evaluation criteria. The performance of each option has to be assessed vis-a-vis the critical threshold condition of that criterion. Thus, fulfilment of the critical threshold values is the first important filter to be passed through in order to qualify for further inspection, e.g. by using a standard multicriteria method. This analysis can be carried out in two ways. The first one is an inspection of a single strategy. The second approach is the comparison of two alternatives or scenarios. In the former procedure, we decide whether the alternative is sustainable or not. In the latter case, by comparing the alternatives, we may decide which choice possibilities scores best when this question is centred around the sustainability issue. This option may be interpreted as a basic form of multicriteria analysis, in which reference values are also included. This approach is different from a benchmarking approach, where the best possible achievement values are used as a frame of reference.

Various applications of the flag model have mainly taken place in land use and environmental studies related to sustainable policy decisions. In this field, the model has shown its capacity to summarise complex problems with acceptable results. The input of the programme consists of the definition of an impact matrix. Thus, each indicator is given its values for each of the considered choice possibilities. Additionally, for each indicator we have to identify the class, the type and the range of its threshold values. The model considers normally quantitative data, but the literature also gives examples of qualitative data (e.g., on sustainable tourism policy).

The flag model can operate both as a classification procedure and as a visualising method. In the former case, for example, in combination with the regime analysis, the flag model can determine the acceptable alternatives according to the examined policy, while next the remaining policy options can then be ranked by the regime method. In the latter case, we can use the flag model to better visualise the results obtained, for example, from the regime method or the rough set procedure to be discussed in the next section.

One of the major aspects of the flag model is its representation module. There are three approaches to the representation: a qualitative, a quantitative and a hybrid approach. The idea of having three possible levels of outcome representation is based upon the necessity for the programmeme to be flexible to the requirements of its users. Rather than to be used as substitutes, the three modes of analysis are complementary to each other.

The qualitative approach only takes into account the colours of the flags. This entails flag counts and cross tabulation. This approach merely displays in various representative ways the results obtained by the evaluation. The quantitative approach defines the values of the indicators that may be acceptable or not. To achieve such results, we need to standardise the indicators which, because they refer to different aspects, are then expressed by different scales of measurement. Finally, the hybrid form regards the existence of both qualitative and quantitative aspects. For example, let us suppose that for a cost indicator CTV = 100, CTV_{max} = 120, and for the three scenarios the indicator values are 114, 119 and 121, respectively. The hybrid form then shows that the first two indicators lead to red flags, while the third indicator is black-flagged (qualitative results). It will also reveal that the outcomes for the second and third indicators are extremely close, while the score for the first is the best (quantitative results).

5. Multidimensional Rough Set Analysis

Another recently developed method for multidimensional classification and assessment problems is rough set analysis. Rough set analysis has been developed within the area of artificial intelligence; its main emphasis is on the question how to define general knowledge and learning processes through deduction mechanisms, and how to differentiate between imprecision and vagueness. In rough set analysis we examine how to draw conclusions from different classes of information (e.g., decisions from imprecise data) and how to determine correlation and relationships among data. In general, through the use of rough set analysis it is possible to recognise cause-effect relationships among the available data, and to underline the importance and the strategic role of specific varieties and the irrelevance of other data (Pawlak 1986, 1991).

The approach focuses on regularities in the data in order to draw inferences and to derive relationships from them which may be less evident, but which can be useful in assessment analysis and policy-making. This approach is mainly based on combinatorial set theory. For this reason rough set analysis overlaps with other mathematical ideas developed to deal with imprecision and vagueness, such as fuzzy logic theory, the theory of evidence, and discriminant analysis. Other comparative analyses have discussed the links among these different mathematical concepts and have pointed out the intrinsic relationships of these methods with rough set analysis (see also Van den Bergh et al. 1997). In recent applications it has become evident that rough set analysis can be applied as a powerful assessment policy method where imprecise information is classified and reduced to determine a coherent policy choice. We will now review rough set analysis in some more detail.

Often the choice among different alternative descriptions of a problem can become very puzzling because of a vague and inaccurate description of the reality we need to examine. Rough set analysis has been designed to reduce the cumbersome character of fuzzy input when we analyse decision situations. More precisely, this approach is used to discover possible cause-effect relationships among the data available, to underline the importance and the strategic role of some data, and to differentiate between irrelevant and relevant data (Pawlak 1986, 1991). The intrinsic attribute of rough set analysis is its ability to manage quantitative as well as qualitative data.

Let us consider a finite universe of objects we would like to examine and classify. For each object we can define a number n of attributes in order to create a significant basis for the required characterisation of the object. If the attribute is quantitative, it will be easy to define its domain. If the attribute is qualitative, we divide its domain into sub-intervals to obtain a more accurate description of the object. We can thus classify our objects in terms of attributes, so that with each object we associate a vector of attributes. The table containing all this organised information will be called the **information table.** From the table of information, we can immediately observe which objects share the same types of attributes. Two objects that are not the same object have an indiscernible relation when they have the same descriptive attributes. Such a binary relation is reflexive, symmetric and transitive.

We can now introduce a fundamental concept in the rough set analysis procedure. Let us imagine that Q is the set of attributes that describe the set of objects U. Let P represent a sub-set of the set of attributes Q, and X represent a sub-set of the set of objects U. We define as a sub-set of $\underline{X}$ those objects which all have the attributes belonging to set P. Such a set is the P-lower approximation of set X, and is denoted as PLX. We then define as P-upper approximation of X, denoted as PUX, the sub-set of U having as its elements all objects belonging to the P set of attributes and which has at least one element in common with set X.

The definition of the upper and lower approximation sets assumes an important role in the rough set methodology. Through these sets we can classify and examine the importance of uncertain information we have collected. Consequently, this approach might lead to an imprecise representation of reality by reducing the information-specific sets. Such an objection against this methodology might be better understood when we recall that the capacity to manipulate uncertain information and the consequent capability of reaching conclusions is one of the most essential assets of the human mind in obtaining knowledge. Therefore, the representation of reality by means of rough set analysis is indeed a reduction of the perceived real phenomena, but it is done in such a way as to enable us to classify, distinguish and express judgements about it.

Until now, we have focused our attention on the classification of uncertain data. Let us now examine the case where we want to express a choice among different alternatives; this is in a pronounced way the case when we are confronted with an assessment problem. We have previously described the information table, and with this table as the information base of an assessment problem, we can identify two classes from the set of attributes: a class of condition attributes and a class of decision attributes.

The class of condition attributes are those which describe the objects following the procedure we have depicted above. The class of decision attributes is defined by all the attributes that the object must have in order to be selected as an acceptable alternative. For instance, a set of objects can be described by values of condition attributes, while classifications of experts are represented by values of decision attributes.

At this point, we must define a decision rule as an implied relation between the description of a condition class and the description of a decision class. The decision rule can be exact or deterministic when the class of decision is contained in the set of conditions, i.e. all the decision attributes belong to the class of the condition attributes. We have an approximate rule when the class of decision attributes is not entirely contained within the set of conditions attributes. Therefore, an exact rule offers a sufficient condition for belonging to a decision class; an approximate rule admits the possibility of this.

The decision rules and the table of information are the basic elements needed to solve multi-attribute choice and ranking problems. The binary preference relations between the decision rules and the description of the objects by means of the condition attributes determine a set of potentially acceptable actions. In order to rank such alternatives, we need to conduct a final binary comparison among the potential actions. This procedure will define the most acceptable action or alternative.

Many applications of the rough set concept have been implemented in various fields of science, e.g., in decision analysis, transport research, urban research and environmental research (see also Capello et al. 1999). One of the most important features of this approach is its capacity to examine quantitative as well as qualitative data. Such data can define vague information and uncertain knowledge that will then be manipulated by the model in the approximation of the data set. Rough set analysis can also be combined with other assessment and evaluation methods. We can consider, for example, applications in a complementary framework of both rough set analysis and regime analysis. With rough set analysis we can determine the classification and approximation of the available information as the basis for the decision process. Then the regime method can elaborate the data according to its assessment rules. A similar application can be conducted with the flag model. Thus, sequential and nested approaches for combining different methods are possible and may expand the range of applications of rough set analysis. In conclusion, rough set analysis corresponds affirmatively to the requirements for a policy assessment methodology as previously defined. Its capacity to examine qualitative and quantitative data, with its main objective to reduce overlapping information and to classify the available data, highlights its overall applicability for comparative research on assessment issues.

6. Comparison of the Three Methods

In this section we will compare the regime, flag and rough set method on the basis of six points of view. Before comparing the three methods, let us summarise some of their properties and limits. Regime analysis is a powerful tool among the assessment methods, since it is able to analyse ordinal as well as cardinal data, and therefore within a multi-objective framework, it can manage a large variety of assessment problems. In the flag model we have shown the possibility of expressing 'fuzzy' and overlapping ranges of critical threshold values for the decision processes, as well as the capacity to represent the results with various devices, thus leading to a user-friendly structure to the programme. Rough set analysis, finally, has the unique quality of being able to synthesise, classify and order the information available to the decision-makers. The three methods can tackle a wide range of assessment problems, but some important questions for future investigation remain. When is one method preferable to another? How can we combine different methods to reach a better result? What kind of results can we achieve? An important consideration is the type of data that each method can analyse (see Table 1).

	Characteristics of Method					
Methods	Mixed Data	Quantitative Data	Definition of Decision Rules	Transparency	Account-ability	User Friendly System
Regime Method	++	++	++	- -	+	−
Flag Model	−	++	+	++	+	++
Rough Set Analysis	+	+	++	−	+	−

Table 1. Comparative study of assessment methodologies

In regional policy decisions (including transportation and environmental policy) the type of data are often qualitative or mixed, i.e. qualitative as well as quantitative. Regime analysis, the flag model and rough set analysis have also a good ability to define proper decision rules. By decision rules we mean the possibility for the decision-maker to identify the type of rules the decision process must fulfill to reach the choice. An example is given by the definition of the weights in the regime analysis or the threshold values in the flag model.

The methods which can define and modify the decision rules have a transparency feature, since the decision-maker can intervene in the assessment process, i.e. in the choice process. Due to this fact, these previous methods can also satisfy the condition of accountability, since the decision-maker, through the determination of the decision rules, will agree and readily support the decision that it has made. The simplicity of the methods is related to the capacity to clarify the assessment process and then allow a friendly use of the method. If we examine the type of results we can obtain, we observe that in particular regime analysis and rough set analysis are able to conduct a full assessment process of choice possibilities.

With these simple elements in mind, it is evident that each assessment tool is chosen in relation to the specific necessity of decision-makers and of the data available to them. Nevertheless, due to the flexibility and compatibility of these three assessment methods, we can interpret them in a compound way where one model may counterbalance the limits of another one. Thus, by considering these three approaches as complementary rather than supplementary, we may achieve more satisfactory results in the assessment process. An example can now be shown for the combined use of regime analysis, the flag model and rough set analysis. Suppose we have to judge a set of alternative transport investment decisions which may destroy some natural areas. We may then ask independent experts to specify critical threshold conditions on biodiversity etc., which should not be surpassed. With the flag model we can then identify the acceptable list of alternatives which satisfies these threshold values. Then through the regime analysis we can define the assessment process of the chosen alternatives by examining the qualitative scores of each of the relevant decision criteria. In this context, policy weights for various criteria can also be introduced. We may simultaneously run a rough set analysis with the complete set of alternatives, i.e. before the selection made by the flag model, in order to compare the consistency of the results and to check whether the final selection fulfils the classification conditions. For the user this is an interesting framework for the analysis, because he is not forced to resort to a single multicriteria method, but he may instead use various methods sequentially or in parallel. This also gives the user greater flexibility and increases the probabilities of obtaining robust results from the assessment procedure. Clearly, combining different methods may incur more costs in terms of time and computer needs, but it may allow us to reach more robust results.

7. Conclusion

In our daily lives we are often confronted with the problem of how to assess choice options and thus how to take decisions in the presence of distinct choice alternatives. Such decisions, however, are often not entirely well-defined and based on rational principles. In particular, assessing policy alternatives is a highly complex process, since it includes -and has to compare- economic, environmental, social, political, and technological aspects. It is principally a communicative process where transparency, simplicity and accountability for the decision-makers are of utmost importance to the success of the decision process. Most evaluation methods implicitly or explicitly use a system of weights, which either represent a policy-maker's perspective or a community's perspective. For example, cost-benefit analysis is based on a collective estimate of all benefits and costs of a policy initiative, whereas the collective interest is defined via the summation of the individual willingness to pay. Thus, this method does not attach importance to individuals or groups not represented in the monetary calculation schemes. For this reason cost-benefit has often been criticised, because costs and benefits may be unevenly distributed.

The **regime analysis** uses explicit weights, either as policy-makers' expressions of importance, or as expert opinions. This method also has a module which calculates the best possible ranking alternatives in case there is no explicit ranking of weights (the principle of 'ignorance').

In the **flag model** there is no explicit weighting, except for the fact that the threshold values reflect some normative expression on acceptability. As mentioned however, the flag model can be extended with a qualitative or quantitative multicriteria analysis, through which weights can be incorporated.

And finally, the **rough set method** does not use explicit weights, but it allows one to assess the importance of moderator variables through statistical techniques. Thus, implicitly it is able to deal with weighting schemes.

It should be added that each of these three methods uses, in one way or another, priority or weighting schemes, sometimes explicitly sometimes implicitly. It is important to realise that the most important step of an assessment procedure is to make the best possible estimate of the expected effects of a policy decision. The mutual weighting of those effects via multicriteria methods is of course an important step, but serves as a sensitivity analysis for the robustness of findings rather than as a 'magical box' from which unexpected results can be obtained.

The above described assessment methods try to cope with the problems of decision situations by trying to define a logical structure based upon rationality and objectivity. Since reality can be defined as a complex system, there are different multi-assessment methods which address the problem of classifying and then making decisions. These methods build upon the principles of cost-benefit analysis, but are also complements and generalisations. Keeping in mind this observation, this paper has reviewed three assessment methods: regime analysis, the flag model, and rough set analysis. These three methods have been chosen, because they give a representative overview on the question of how to approach a multi-objective assessment problem. In a decision situation we encounter various obstacles, such as the characterisation of alternatives, or the definition of the relative weights among the potential decisions. In this context, these three methods can operate separately according to the type of 'obstacle' we need to overcome, but they can also operate in a sequential way. By this we mean that certain problems can be better solved by a specific method, and then the assessment problem can be carried out with another approach. Therefore, these three approaches may be thought of as complementary to traditional project methods as well as to each other. Altogether they offer a solid portfolio of applicable assessment methods for strategic policy analysis.

The conclusions from this paper are rather straightfonvard. There is an urgent need as well as a great potential for the application of systematic assessment methods for strategic policy analysis. Such methods aim to evaluate the pros and cons of a planned policy initiative. The foundation of such methods rests on conventional cost-benefit analysis, but in the light of the often incommensurable and qualitative aspects of transport decisions, there is now more scope for complementary, adjusted evaluation techniques based on multicriteria or multi-assessment methods. There is clearly no single assessment method which can satisfactorily and unequivocally evaluate all complex aspects of modern policy. The choice of assessment methods in any given policy context therefore depends on the features of the policy problem at hand, on the aims of the policy analysis, and on the underlying information base. This will ensure coherence between the assessment method used and the actual choice problem to be tackled.

REFERENCES

Ackoff, R.L. (1981). *Creating the Corporate Future: Plan or Be Plannedfor.* Wiley, New York.

Bergh, J.C.J.M. van den, K.J. Button, P. Nijkamp and G.C. Pepping (1997). *Meta-Analysis in Environmental Economics.* Kluwer, Boston.

Capello, R, P. Nijkamp and G. Pepping (1999). *Sustainable Cities and Energy Policies.* Springer-Verlag, Berlin.

Hermanides, G. and P. Nijkamp (1998). "Multicriteria Evaluation of Sustainable Agricultural Land Use: a Case Study of Lesvos", *in Multicriteria Analysis for Land-Use Management.* E. Beinat and P. Nijkamp (eds.), Kluwer Academic Publishers, Dordrecht, Dordrecht, pp. 61-78.

Jackson, P. *(1990). Introduction to Expert Systems.* Addison-Wesley, New York.

Janssen, R. (1991). *Multiobjective Decision Support for Environmental Problems.* Ph.D. Dissertation, Free University of Amsterdam, Amsterdam.

Kacprzyk, J. and R.R. Yager (1990). "Using Fuzzy Logic with Linguistic Quantifiers in Multiobjective Decision-Making and Optimization: a Step Towards More Human-Consistent Models", in *Stochastic versus Fuzzy Approaches to Multiobjective Mathematical Programming under Uncertainty.* R. Slowinski and J. Teghem (eds.), Kluwer Academic Publishers, Dordrecht.

Leung, *Y.* (1997). *Intelligent Spatial Decision Support Systems.* Springer-Verlag, Berlin.

Mintzberg, H. (1979). *The Structuring of Organisations: a Synthesis of the Research.* Prentice Hall. New York.

Nijkamp, P. and E. Blaas (1995). *Impact Assessment and Evaluation in Transportation Planning.* Kluwer, Dordrecht.

Nijkamp, P. and G. Pepping (1998). "Meta-Analysis for Explaining the Variance in Public Transport Demand Elasticities in Europe". *Journal of Transportation and Statistics,* vol. 1, pp. 1-14.

Nijkamp, P., P. Rietveld, and H. Voogd (1992). *Multicriteria Evaluation in Physical Planning.* Elsevier, Amsterdam.

Nijkamp, P. and H.J. Scholten (1993). "Spatial Information Systems: Design, Modelling and Use in Planning". *International Journal of Geographical Information Systems.* vol. 7, pp.85-96.

OECD (1993). *Corps Central d 'Indicateurs de l 'OECD pour les Exames des Performances Environmentales.* Paris.

Ouwersloot, J., and P. Nijkamp (1998), Multidimensional Sustainability Analysis: The Flag Model, *Theory and Implementation of Economic Models for Sustainable Development* (J.C.J.M. van den Bergh and M.W. Hofkes, eds.), Edward Elgar, Cheltenham, 1998, pp. 255-273.

Pawlak, Z. (1986). "On Learning - A Rough Set Approach". *Lecture Notes in Computer Science.* Springer-Verlag. vol. 208, pp. 197-227.

Pawlak, Z. (1991). *Rough Sets: Theoretical Aspects of Reasoning about Data.* Kluwer Academic Publishers. Dordrecht.

Rietveld, P. and F. Bruinsma (1998). *Are Investments in Transport Infrastructure Effective?* Springer-Verlag, Berlin.

Sikow-Magny C. and E. Niskanen (1998). "Decision-Making Criteria for National Economic Policy, Transportation Policy and Road Policy Levels". Paper presented at the 8" World Conference on Transport Research, 12-17 July, Antwerp.

Simon, H. *(1960). The New Science of Management Decision.* Harper and Row, New York.